文章编号：1007-2322（2021）00-0001-10　　文献标志码：A　　中图分类号：TM714

一种改进 Transformer 的电力负荷预测方法

黄飞虎[1,2]，赵红磊[1]，弋沛玉[2]，李沛东[2]，彭舰[2]
（1. 四川中电启明星信息技术有限公司， 四川成都 610000；2. 四川大学计算机学院，四川成都 610065）

An Improved Power Load Forecasting Method Based on Transformer

HUANG Feihu[1], ZHAO Honglei[1], YI Peiyu[2], LI Peidong[2], PENG Jian[2]
（1. Aostar Information Technologies Co., Ltd., Chengdu 610000, Sichuan Province, China
2. College of Computer Science, Sichuan University, Chengdu 610065, Sichuan Province, China）

摘要：[1]负荷预测是电网系统中很多应用的关键部分，具有重要作用。然而，由于电网负荷的非线性、时变性和不确定性，使得准确预测负荷具有一定的挑战。充分挖掘负荷序列的潜在特征是提升预测准确率的关键。本文认为在特征提取时应该充分利用负荷序列的位置信息、趋势性、周期性和时间信息，同时还应构建更深层次的神经网络框架进行特征挖掘。因此，本文提出了基于特征嵌入和 Transformer 框架的负荷预测模型，该模型由特征嵌入层，Transformer 层和预测层组成。在特征嵌入层，模型首先对历史负荷的位置信息、趋势性、周期性和时间信息进行特征嵌入，然后再与天气信息进行融合，得到特征向量。Transformer 层则接受历史序列的特征向量并挖掘序列的非线性时序依赖关系。预测层通过全连接网络实现负荷预测。从实验结果来看，本文模型的预测性能优于对比模型，体现了该模型的可行性和有效性。

关键词：负荷预测；特征嵌入；Transformer 框架；神经网络

Abstract：Load forecasting is a key part of many applications in power grid systems and plays an important role. However, due to the non-linearity, time-varying and uncertainty of the grid load, it is challenging to accurately predict the load. Mining the potential characteristics of load sequence plays a key role to improve the accuracy of forecasting. This paper considered that the position information, trend, periodicity and time information of the load sequence should be fully utilized 2017YFB0202403 in feature extraction, and a deeper neural network framework should be constructed for feature mining. Therefore, a load forecasting model based on feature extraction and Transformer framework was proposed in this paper. The model consists of a feature embedding layer, a Transformer layer and a prediction layer. In the feature embedding layer, the location information, trend, periodicity and time information of the historical load are embedded into a characteristic vector. Then it outputs feature vector after fusing the obtained vector with meteorology information. Transformer layer aims to mining the temporal nonlinear dependence hidden in the sequence based on the obtained feature vectors of load sequence. The prediction layer does the load forecasting task through a fully connected network. According to the experimental results, the proposed model in this paper outperforms the baselines, which verifies the feasibility and effectiveness of the model.

Keywords: Load forecasting; Feature embedding; Transformer framework; Neural network

DOI：10.19725/j.cnki.1007-2322.2021.XXXX

0　引言

负荷预测是电网系统中很多应用场景的重要环节。一方面，用户侧的电力需求是动态变化的，由于电能的特点是即发即用，因此负荷供需的动态平衡需要负荷预测的支持，才能保障整个电力系统稳定、高效地运行；另一方面，国家提出了构建以新能源为主体的新型电力系统的目标，未来新能源（风、光等）将成为电网系统的重要部分，然而新能源的出力具有不确定性[1]，如何保障电网的稳定运行需要准确预测用电负荷，合理制定发电计划；此外，电动汽车的普及和电力市场推进的背景下[2]，电网负荷的波动性进一步增强。可见，构建具有准确预测能力的负荷预测模型具有迫切性和实际应用价值。

负荷预测任务的难点在于其受诸多外部因素

基金项目：国家重点研发计划项目(2017YFB0202403);四川省科技计划项目(2018GZDZX0010, 2020YFG0308, 2017GZDZX0003)
基金英文：Project Supported by National Key Research and Development Program of China (2017YFB0202403). Sichuan Science and Technology Program (2018GZDZX0010, 2020YFG0308, 2017GZDZX0003);

的影响[1,3]。比如，天气因素和市场因素会给短期、超短期的负荷带来波动，政府政策因素、居民收入因素会给长期、中长期负荷带来影响。此外，数据获取也是影响负荷预测精度的重要因素。影响负荷的外在因素太多，很多因素难以获取。这种情况下，很难有预测模型能够反映真实的负荷与因素之间的关系，从而影响模型预测效果。

负荷预测任务主要解决的问题是电网负荷的非线性、时变性和不确定性。非线性体现了负荷与影响因素之间的复杂函数关系。时变性则反映的是负荷自身的相关性，即负荷的当前状态与过去的历史状态具有相关性，可以体现为时序上的趋势性、季节性等。不确定性则表现为很多难以观测的因素对负荷变化产生的影响。

现有负荷预测方法可以分为：基于数学模型的预测方法、基于传统机器学习的预测方法和基于深度学习的预测方法[2-3]。

（1）基于数学模型的预测方法，常见的有自回归移动平均(Auto Regression Moving Average, ARMA)[4]、自回归(Auto Regression, AR)[5]等时间序列模型。此外，也有基于卡尔曼滤波的预测方法[6-7]，基于指数平滑度预测方法[8]。这类方法简单，运行速度快，能够解决负荷的时变性。但是不能很好地解决非线性和不确定性问题。

（2）基于传统机器学习的预测方法。机器学习算法可以有效处理非线性问题[9]。这类方法常用的有支持向量机（Support Vector Machines, SVM）、决策树、多层感知机（Multilayer Perceptron，MLP）等。比如，文献[10]结合数据挖掘预处理技术和SVM 对负荷作预测，相比单纯的 SVM，减少了数据量，提高了预测的速度和精度。文献[11]基于相似数据选取和梯度提升决策树对负荷进行预测。文献[12]提出了一种基于改进支持向量机的混合预测模型，并利用粒子群算法作参数优化。文献[13]基于多层感知机构建混合模型实现负荷预测，模型中数据由梯度增强机进行特征学习，并传入 MLP 作为特征输入。基于机器学习的方法可以实现负荷预测，但是面对复杂的电力系统，其对数据特征的挖掘还是略显不足。

（3）基于深度学习的预测方法。随着计算机性能的增强，深度学习技术因其强大的非线性映射和自适应能力在各个行业都有应用且效果都不错。比如，以长短期记忆网络 LSTM（Long Short-Term Memory, LSTM）和门控循环单元（Gate Recurrent Unit, GRU）为主的循环神经网络（Recurrent Neural Network, RNN）在负荷预测中有广泛应用。比如，文献[14]基于长短期记忆网络提出了一种针对区域级负荷的深度超短期预测方法。也有很多学者将循环神经网络与其他方法相结合实现负荷预测。比如，文献[15]中，作者认为电力系统相关数据很多且是非线性的，利用皮尔逊相关系数并不能很好地挖掘数据特征，则提出利用最大互信息系数对多源特征进行选择，然后结合 LSSVR、LSTM 和 XGBoost 进行模型构建。为了充分利用深度学习模型非线性映射和自适应的优势，学者们通常将其与很多特征提取方法相结合。比如，文献[16]中，作者利用经验模态分解方法将负荷数据进行分解，然后基于 CNN 深度分析了负荷数据在不同频率下的时序特征，最后由 LSTM 利用学到的特征进行预测。文献[17]利用小波分解法对负荷数据进行预处理，然后结合神经网络做负荷预测。Transformer 模型[18]是谷歌在 2017 年提出的，该模型的最大特点为它抛弃了 RNN 和 CNN 等网络结构。Transformer 模型最初在机器翻译领域大显生手，最近很多学者将其运用到了序列数据预测、目标检测等领域，均取得了不错的效果[19]。比如，文献[20]构建了基于注意力的时空图网络模型作交通流量预测，其注意力的实现就是利用 transformer 模型。文献[21]则是利用 Transformer 构建了时空特征提取模块，不同的是其只采用了 Transformer 的编码块。

针对负荷预测的难点，本文将特征提取技术和深度学习结合，提出了基于改进 Transformer 的负荷预测模型。相比现有工作，本文模型的在 Transformer 编码块和预测层进行了改进。相比原始 Transformer 模型直接对输入数据进行编码，本文构建了特征嵌入层对输入的历史负荷进行特征嵌入。在预测层，输入数据为过去 h’（小于 h）时刻的特征，经解码块之后，将得到的输出特征直接传入预测层实现多步预测。相比原始 Transformer 模型中，采用 t-1 时刻的预测值作为输入（迭代预测），本文方法（直接预测）可以避免误差累积。特别是做多步预测任务时，更能有效提升预测性能。

1 算法模型原理

1.1 注意力机制

注意力机制最早在机器视觉领域被提出，其来源是借鉴人类的注意力。抽象到机器学习领域，就是通过训练让模型对输入数据分配不同的权重。其中，这些权重就是注意力[22]。

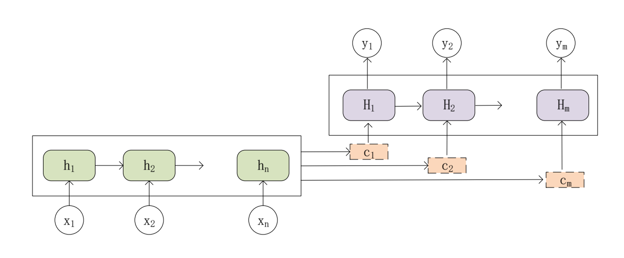


图 1 带注意力机制的编码-解码框架

Fig. 1 Encoder-decoder framework with attention

常见的注意力机制是软注意力机制（Soft Attention，SA）。注意力机制可以用于任何模型，本文以编码-解码框架为基础，对 SA 作介绍。编码-解码框架在自然语言处理、时间序列预测中有广泛应用。如图 1 所示，输入序列为$[x_1,x_2,\cdots,x_n]$，输出序列为$[y_1,y_2,\cdots,y_m]$，其中注意力结果 c_i 的计算公式为：

$$c_i=\sum_{j=1}^{n} a_{ij}h_j \tag{1}$$

其中 n 是编码层的输入数据长度，h_j 则是编码层中第 j 个输入数据的隐藏层状态。a_{ij} 表示解码层输出第 i 个值时编码层第 j 个数据的注意力分配系数，计算方法为：

$$a_{ij}=F(\boldsymbol{H}_i,h_j) \tag{2}$$

$\boldsymbol{H}_i$ 是解码层中第 i 个数据的隐藏层状态。F 为函数，计算 $\boldsymbol{H}_i$ 和 h_j 的相似度。然后函数 F 的输出经过 softmax 进行归一化就得到了注意力分配概率分布。

从公式（1）和（2）可以将注意力机制用更一般化的方法表示：

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\sum_{j=1}^{n} Softmax(Similarity(\boldsymbol{Q},\boldsymbol{K}_j))*\boldsymbol{V}_j \tag{3}$$

其中，$\boldsymbol{Q}$ 称为查询矩阵，$\boldsymbol{K}$ 为关键字矩阵，$\boldsymbol{V}$ 为值矩阵。

除了 SA，还有硬注意力机制（Hard Attention）、全局注意力机制（Global Attention）、局部注意力机制（Local Attention）和自注意力机制（Self-Attention）[22]。硬注意力机制中 a_{ij} 的取值为 0 或 1，全局注意力机制是考虑编码层中所有的隐藏层状态，局部注意力机制则是考虑部分隐藏层状态。自注意力机制是由谷歌在 Transformer 模型中提出。自注意力机制中 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 均属于同一对象，其他注意力机制中 $\boldsymbol{Q}$ 属于一个对象，$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 属于另一个对象。

1.2 Transformer 模型

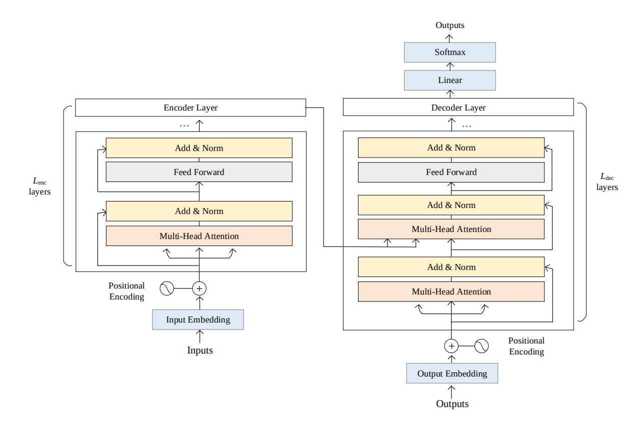


图 2 Transformer 模型框架

Fig. 2 Framework of transformer model

Transformer 模型由编码块和解码块组成，如图 2 所示。编码块是由 L_{enc} 个独立的编码层堆叠而成。每一个编码层里面包含多头注意力层、全连接层和正则化层。同理，解码块也是 L_{dec} 个解码层堆叠而成。解码层与编码层的区别在于，每一个解码层有两个多头注意力层。

先介绍解码层的多头注意力，其结果可以表示为：

$$Multihead(H)=concat(head_1,\ldots,head_u)\boldsymbol{W}^O \tag{4}$$

其计算过程为，首先将 u 个注意力表示进行拼接，然后与 $\boldsymbol{W}^O$ 作矩阵乘法。结合公式（3），单个注意力块可视为 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 的函数，即：

$$head_i=softmax(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\boldsymbol{V}) \tag{5}$$

其中，$\boldsymbol{Q}\in R^{n\times d_k}$，$\boldsymbol{K}\in R^{n\times d_k}$，$\boldsymbol{V}\in R^{n\times d_v}$。$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 是由输入数据进行编码之后再次作线性映射得到：

$$\boldsymbol{Q}=\tilde{\boldsymbol{X}}\boldsymbol{W}^Q \tag{6}$$

$$\boldsymbol{K}=\tilde{\boldsymbol{X}}\boldsymbol{W}^K \tag{7}$$

$$\boldsymbol{V}=\tilde{\boldsymbol{X}}\boldsymbol{W}^V \tag{8}$$

这里 $\boldsymbol{W}^Q$, $\boldsymbol{W}^K$,$\boldsymbol{W}^V$ 是可学习的参数。$\tilde{\boldsymbol{X}}$是输入数据与位置编码相结合得到的特征矩阵。给定 n 个输入数据，每个输入项$\boldsymbol{X}_t\in R^{1\times d}$为 d 维向量，$\tilde{\boldsymbol{X}}_t$定义为：

$$\widetilde{X}_t=X_t+e_t \tag{9}$$

e_t是位置编码函数，其计算公式为：

$$e_t=\begin{cases} sin\left(\dfrac{t}{10000^{\frac{2i}{d_m}}}\right) & \text{if } t=0,2,4,\ \dots \\ cos\left(\dfrac{t}{10000^{\frac{2i}{d_m}}}\right) & \text{if } t=1,3,5,\ \dots \end{cases} \tag{10}$$

对于解码层，因其有两个多头注意力层，其中第一个注意力层与解码层的一样。第二个注意力层的 $\boldsymbol{K}$ 和 $\boldsymbol{V}$ 是解码块的输出，$\boldsymbol{Q}$ 是正则化层的输出。*Transformer* 中正则化层结构都一致，主要由残差连接和正则化操作构成：

$$norm_{cur}=Normalization(\mathbf{z},norm_{pre}) \tag{11}$$

其中 $\mathbf{z}$ 是注意力层或者全连接层的输出。

2 本文模型

负荷预测的任务为，给定获取 h 时刻的历史负荷序列$[\boldsymbol{X}_{t-h+1},\dots,\boldsymbol{X}_t]$，学习函数 f 对未来 t+1 时刻的负荷值做预测，其形式化的定义为：

$$\boldsymbol{X}^{t+1}=f([\boldsymbol{X}^{t-h+1},\dots,\boldsymbol{X}^{t}]) \tag{12}$$

为了实现具有良好预测性能的负荷预测模型，本文认为需要解决两个问题，即特征提取和非线性依赖。特征提取，是基于历史数据和额外的天气数据进行特征挖掘为模型提供输入特征。由于负荷数据的影响因素很多，因此尽可能地挖掘丰富的特征信息是特征提取阶段的目的。非线性依赖是基于输入特征数据挖掘其潜在的在时间维度的非线性依赖关系，其目的是尽可能地挖掘负荷的变化模式。在模型中，本文设计了特征提取层解决特征提取问题，基于 Transformer 模型解决非线性依赖问题，最后通过预测层实现负荷预测。图 3 为本文预测模型的框架。

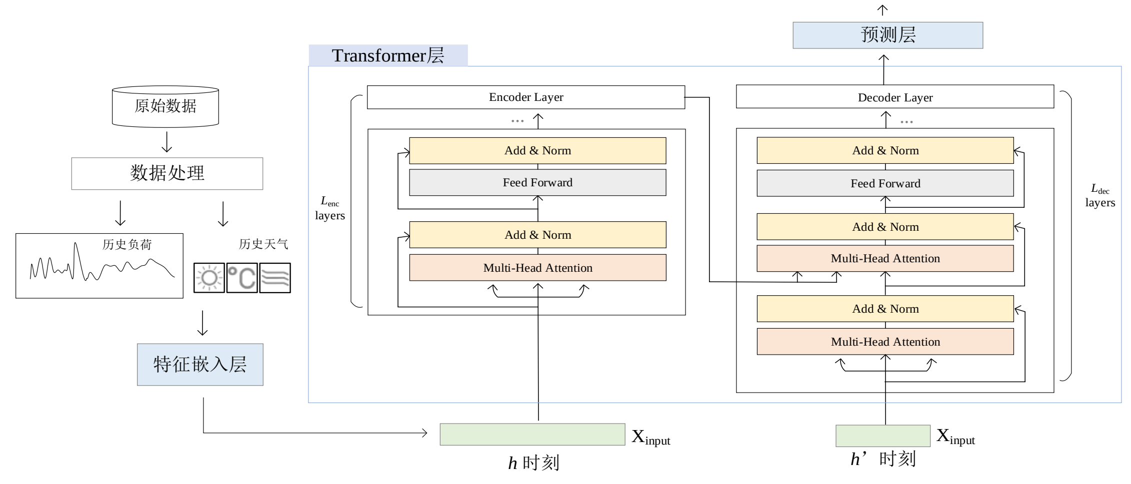


图 3 预测模型框架

Fig. 3 Framework of Prediction model

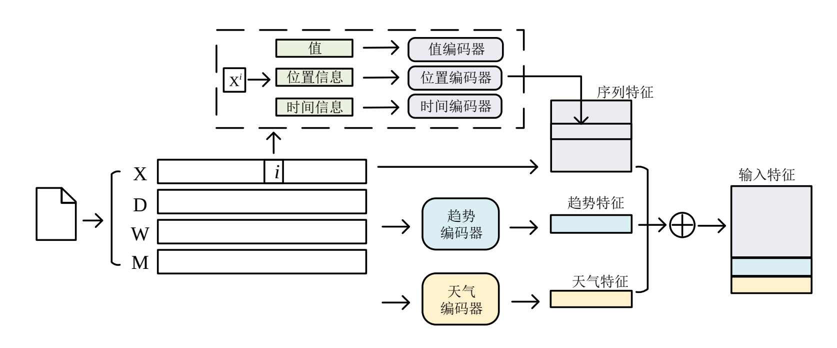


图 4 特征嵌入层框架

Fig. 4 Framework of feature embedding layer

2.1　特征嵌入层

特征嵌入层的主要目的是为下一层提供丰富的特征信息。因此，本文在特征嵌入层从三个方面进行了特征提取和融合，分别为序列特征、趋势特征和天气特征。如图 4 所示，在该层本文分别设计了不同的编码器用于对这三类特征进行嵌入。

从历史负荷数据中提取长度为 h 时刻负荷数据，可以得到负荷序列$[\boldsymbol{X}_{t-h+1},\ldots, \boldsymbol{X}_t]$。在已有文献中，大多数算法是直接对序列 $\boldsymbol{X}$ 进行特征挖掘，比如文献[9,23]。与这些模型类似，本文也采用了值编码器 E_v 对其值进行特征映射。同时，本文认为负荷序列 $\boldsymbol{X}$ 中的时间信息、位置信息对于下一层的非线性依赖关系挖掘都有重要作用。因此，对于位置信息，本文采用位置编码器 E_p（公式 10）对 $\boldsymbol{X}$ 中每一项 $\boldsymbol{X}_i$ 进行位置编码。其位置信息即是该项 $\boldsymbol{X}_i$ 在当前序列中的顺序，这里即为 i。该方法与传统 Transformer 模型中的位置编码方法一样。引入位置编码是让编码块中的注意力机制能够对序列输入的数据进行定位。对于时间信息，本文采用时间编码器 E_t 对当前时间进行编码，该模块有助于在数据中引入时间信息。因为用户的用电需求与时间具有很强的相关性。对于每一项 $\boldsymbol{X}_i$，首先提取其时间信息（月份、日、周、时、分信息），假定其当前时间为’2021-03-03 15:30:30’，则提取的信息包含：3 月份、3 号、星期三、下午 15 点、30 分。用数值向量表示为[3,3,3,15,30]。对于分钟信息，在实际处理时，以 15 分为一个时间段，即把 1 小时分成 4 段。因此，在输入时间编码器时，该时间信息应该表示为[3,3,3,15,2]。模型中，值编码器和时间编码器为简单的全连接网络。位置编码器直接采用公式 10 进行计算，则不需要参数。用公式形式化地表示序列 $\boldsymbol{X}$ 的处理过程为：

$$\boldsymbol{X}_{seq}=E_v(\boldsymbol{X})+E_p(\boldsymbol{X})+E_t(\boldsymbol{X}) \tag{13}$$

在数据预处理时，除了从历史负荷数据中提取长度为 h 时刻负荷序列$[\boldsymbol{X}_{t-h+1},\ldots, \boldsymbol{X}_t]$，还需要获取过去 h 天在 t 时刻的负荷数据构建 $\boldsymbol{D}$，过去 h 周在 t 时刻的负荷数据构建 $\boldsymbol{W}$，以及过去 h 月在 t 时刻的负荷数据构建 $\boldsymbol{M}$。本文将 $\boldsymbol{D}$、$\boldsymbol{W}$ 和 $\boldsymbol{M}$ 定义为数据的趋势信息。这些信息包含了过去相对来说较长一段时间的负荷变化趋势，能够反映一定的周期性和趋势性。在模型中，本文将 $\boldsymbol{D}$、$\boldsymbol{W}$ 和 $\boldsymbol{M}$ 进行叠加构成趋势矩阵$\boldsymbol{T}\in R^{3\times h}$，然后利用卷积操进行特征提取：

$$\boldsymbol{X}_{tre}=Conv1d(\boldsymbol{T}) \tag{14}$$

其中 $Conv1d$ 为一维卷积操作。

由于天气与用户用电也有很大相关性，因此在模型中还引入了天气数据进行特征融合。由于天气数据都是数值类型，则可将天气数据构成数值向量 $\boldsymbol{Q}$，作为全连接网络的输入，即：

$$\boldsymbol{X}_{wea}=E_w(\boldsymbol{Q}) \tag{15}$$

最后，将这些特征进行融合，即：

$$\boldsymbol{X}_{input}=concat(\boldsymbol{X}_{seq},\boldsymbol{X}_{tre},\boldsymbol{X}_{wea}) \tag{16}$$

2.2 Transformer 层

针对负荷数据的非线性和时变性特征，本文采用 Transformer 模型进行挖掘。相比常用的 LSTM 和 GRU 网络，Transformer 模型更适合挖掘序列中的时序依赖关系。LSTM 和 GRU 的缺点在于，其编码阶段只有最多一个单元的隐藏特征可以传递到解码层，这就导致部分信息丢失[22]。对于负荷数据这类波动性较大的任务来说，应该充分利用前面提取的隐藏特征。其实，在时序数据预测任务中，对于 LSTM 和 GRU 的这种部分信息丢失的情况，学者们提出了可以通过注意力机制解决。Transformer 模型的特点在于，它不仅引入了注意力机制，而且还采用了更深层的网络。这些技术使得 Transformer 在时序依赖关系挖掘方面比 LSTM 和 GRU 更加擅长。因此，在模型中本文将特征嵌入层得到的特征$\boldsymbol{X}_{input}$输入 Transformer 层，则可得到具有时序依赖的特征信息$\boldsymbol{X}_{dep}$，其计算公式为：

$$\boldsymbol{X}_{dep}=Transformer(\boldsymbol{X}_{input}) \tag{17}$$

2.3 预测层

通过 Transformer 层获取到特征信息$\boldsymbol{X}_{dep}$后，便可利用预测层实现负荷预测。本文采用的预测层网络框架如图 5 所示，该网络由 1 个全连接层构成。这里需要强调的是，Transformer 架构是一个序列到序列的模式，在解码时需要有输入数据。在传统的 Transformer 模型中，解码块的输入数据是 t-1 时刻的预测值。采用这种方式的缺点在于，预测误差会进行累积。文献[24]认为，时间序列预测可以采取直接预测的方式，防止误差累积。本文在设计 Transformer 解码块时也采用这种方式，其只需要一次输入，然后进行特征学习，解码块的输出特征则

直接输入预测层实现负荷预测。解码块的输入为过去 h'（小于 h）时刻的特征。

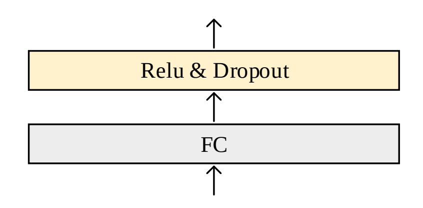


图 5 预测层框架

Fig. 5 Framework of prediction layer

3 实例分析

3.1 数据集与评价指标

本文采用的数据集[2]源于美国能源信息署的公开数据平台，本文下载了纽约市 2015 年 7 月至 2021 年 4 月的数据。数据间隔时间以小时为单位。

采用评价方法有平均绝对百分比误差（Mean Absolute Percentage Error，MAPE）和均方根误差差（Root Mean Squared Error，RMSE）作为算法的评价指标，计算公式如下：

$$\begin{cases} MAPE=\frac{100\%}{n}\sum_{i=1}^{n}|x_i-\hat{x}_i| \\ RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{x}_i)^2} \end{cases} \quad (18)$$

公式中，n 为样本数，x_i为预测负荷值，$\hat{x}_i$为实际负荷值。

3.2 对比算法

（1）HA，即历史平均，该方法是时序数据预测经常采用的对比方法，其主要特点是可以感知序列的趋势特征。实验中，输入数据长度为 h。

（2）ARIMA，该方法是经典的时间序列方法，在负荷预测任务中也常作为对比算法。实验中，采用模型参数为(0,0,1)。

（2）GBRT，即渐进梯度回归树（Gradient Boost Regression Tree），可处理不同类型的数据，具有很强的预测能力。该方法也是时序数据预测经常采用的对比方法。实验中，利用 sklearn 提供的函数实现，主要参数为估计器数量 500，最大深度 3。

（3）LSTM，即长短期记忆网络，是循环神经网络的经典网络，能够挖掘序列数据的非线性时间依赖关系。实验中，采用 2 层 LSTM，输出维度为 64。

（4）BP，即反向传播(Back Propagation)神经网络，是经典的神经网络，在时序预测和负荷预测任务中有广泛应用。实验中，隐藏层维度为 32，输出维度 64。

（5）CECL[16]，即基于聚类经验模态分解的模型，文中作者将其与 CNN、LSTM 结合，主要用于对比本文模型采用的特征提取层和 Transformer 层。实验中，序列分解为 17 个子序列，聚类数为 10，CNN 采用 3x3 的卷积核，LSTM 的输出维度为 64，层数为 1。

3.3 参数设置

在特征提取层，E_v、E_t 和 E_w 三个编码器是全连接网络，则参数为全连接网络的输出，本文将这三个全连接网络的输出均设为 d_e。在 Transformer 层的参数分别有为 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 的输出维度 d_k, $\boldsymbol{V}$ 的输出维度 d_v, 注意力头数 head, 编码块层的层数 L_{enc}, 解码层的层数 L_{dec}，编码块输入历史数据长度 h，解码块输入历史数据的长度 h'。本文将 d_k 的维度与 d_e 保持一致。在预测层，全连接网络的输出为预测值，维度为 1。在预测层，模型是将解码层的输出按行展开，则其维度与编码层 h'以及 d_v、head 相关。因此，在预测层，需要设置的参数为 h'。在模型参数选择时，对参数组合不同的值，并采取网格搜索的方式确定模型参数。d_e、d_k 和 d_v 的取值集合设为[8，16，32，64]。head、L_{enc} 和 L_{dec} 的取值集合为[1，2，3]。h 设置为 168，h'设置为 48。这里的预测任务均是间隔时间为 1 小时的负荷序列数据。

表 1 不同参数组合及其预测效果

Table. 1 Different parameter setting and its prediction performance

参数					评价指标	
d_e/d_k	d_v	head	L_{enc}	L_{dec}	RMSE/kWh	MPAE/%
8	8	1	1	1	581.32	11.01
8	16	1	1	1	578.45	10.31
16	32	1	1	1	551.61	10.06
32	64	1	1	1	544.25	9.88
32	64	2	1	1	521.09	9.76
32	**64**	**3**	**2**	**2**	**497.54**	**9.42**
32	64	3	3	3	517.31	9.51

表 1 给出了在不同参数组合情况下，模型在数据集上的预测效果。从实验结果来看，组合[32，64，3，2，2]效果最好，因此将其作为模型参数。此外，

数据集：https://www.eia.gov/electricity/gridmonitor/dashboard/electric_overview/US48/US48

在模型训练阶段，采用 Adam 算法，学习率设置为 10^{-4}，batch 大小设置为 100，训练次数为 30。训练目标函数为：

$$\text{Loss}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{x}_i)^2 \tag{18}$$

n 为样本数，x_i为预测负荷值，$\hat{x}_i$为实际负荷值。

3.4 预测结果分析

3.4.1 负荷预测效果对比

表 2 模型预测效果对比

Table. 2 Comparison of model prediction performance on the first datasets

模型	RMSE/kWh	MAPE/%
ARIMA	1656.02	21.08
LSTM	840.16	13.71
BP	778.37	13.79
HA	1748.04	22.17
GBRT	717.48	12.02
CECL	778.25	13.28
Transformer	629.51	10.06
本文模型	**497.54**	**9.42**

表 2 为各模型在数据集 1 上的负荷预测结果，因时间间隔为 1 天，则称其为日负荷预测。从实验结果来看，本文模型的预测效果最好，其次是 Transformer 模型。对比模型中，ARIMA 和 GBRT 均是时间序列预测的经典模型。然而，ARIMA 在该数据集上表现不好，GBRT 是采用集成学习思路实现预测，效果不错。CECL 通过序列分解挖掘序列特征也有不错的预测效果。HA 是通过对历史负荷取均值实现预测，获取信息有限，在数据集上表现不好。图 6 给出了各模型的预测实例，时间间隔为 1 天。与真实符合相比，可以看出本文模型的具有良好的预测性能。

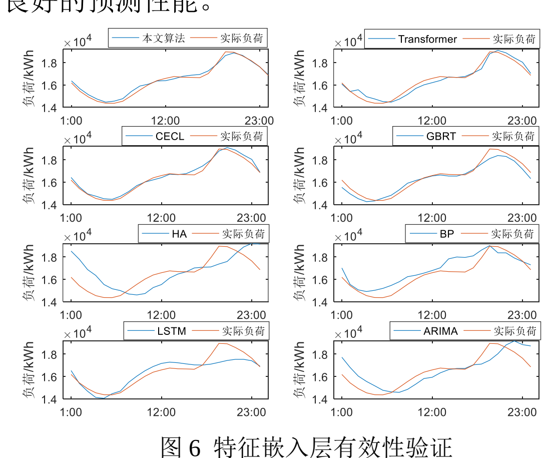


图 6 特征嵌入层有效性验证

Fig. 6 Validation of feature embedding layer

3.4.2 特征嵌入层有效性验证

为了提高模型负荷预测效果，本文模型设计了特征提取层，分别将天气信息、趋势信息以及时间信息融合到模型中。在实验中，本文分别对比了去掉这些信息时模型的预测效果，以验证这些信息对于模型的有效性，如图 7 所示。可以看出，当去掉某些特征信息之后，模型的预测效果都会受到影响。特别是时间信息去掉之后，模型的预测效果下降最多。这表明时间信息对于模型预测具有很重要的作用。这也从另一方面验证了，用户用电与时间具有相关性。

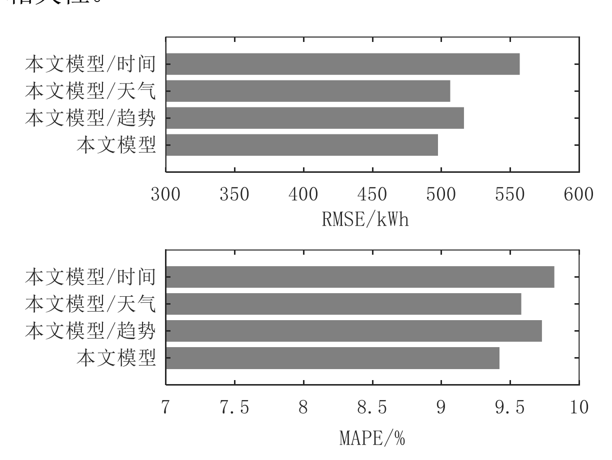


图 7 特征嵌入层有效性验证

Fig. 7 Validation of feature embedding layer

3.4.3 预测层有效性验证

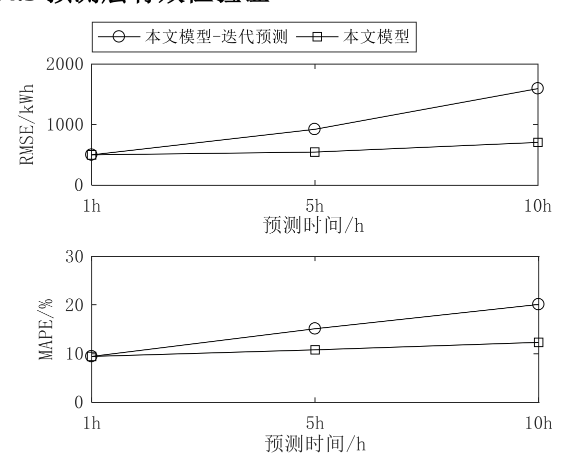


图 8 预测层有效性验证

Fig. 8 Validation of prediction layer

为了验证本文模型在解码块做的改进（即由原始 Transformer 的迭代预测改为直接预测）是否有效，图 8 给出了两种预测方式在未来 1 天、未来 5 天、10 天的预测结果。实验从结果来看，本文模型随着预测时间的增加，模型预测性能并未大幅度下

降。相反，迭代预测在多步预测中由于误差的累计，随着预测步长的增加预测性能下降较快。

4 结论

针对负荷预测任务，本文提出了基于特征嵌入和 Transformer 框架的预测模型。模型中，设计了特征嵌入层，充分利用负荷序列的位置信息、时间信息和趋势信息进行特征编码，这是本文模型的主要创新点之一。同时，在预测层采用直接预测方式，将解码块的输出特征直接传入预测层实现多步预测，能够减少误差累积，这也是本文模型的主要创新点之一。在实验中，通过真实负荷数据集，验证了验证本文模型的有效性和可行性。

参考文献

[1] 王栋.电力系统负荷预测综述[J].电气开关,2020,58(01):6-8+20.
WANG Dong. The Summary of Power System Load Forecasting[J]. Electric Switchgear, 2020,58(01):6-8+20.

[2] 杨智宇,刘俊勇,刘友波,等. 基于自适应深度信念网络的变电站负荷预测[J]. 中国电机工程学报,2019,39(14):4049-4061.
YANG Zhiyu, LIU Junyong, LIU Youbo, et al. Transformer load forecasting based on adaptive deep belief network[J]. Proceedings of the CSEE, 2019,39(14):4049-4061.

[3] 朱俊丞,杨之乐,郭媛君,等. 深度学习在电力负荷预测中的应用综述[J]. 郑州大学学报(工学版),2019,40(05):13-22.
ZHU Juncheng, YANG Zhile, GUO Yuanjun, et al. Deep learning applications in power system load forecasting: a survey[J]. Journal of Zhengzhou University(Engineering Science), 2019,40(05):13-22.

[4] Chen J F，Wang Weiming，Huang Chaoming．Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting[J]. Electric Power Systems Research，1995，34(3)：187-196．

[5] Hagan M T，Behr S M．The time series approach to short term load forecasting[J]．IEEE Transactions on Power Systems，1987，2(3)：785-791．

[6] 赵峰，孙波，张承慧．基于多变量相空间重构和卡尔曼滤波的冷热电联供系统负荷预测方法[J].中国电机工程学报，2016, 36(2): 399-406．
ZHAO Feng，SUN Bo，ZHANG Chenghui．Cooling, Heating and Electrical Load Forecasting Method for CCHP System Based on Multivariate Phase Space Reconstruction and Kalman Filter [J]．Proceedings of the CSEE, 2016，36(2)：399-406.

[7] 陈培垠，方彦军．基于卡尔曼滤波预测节假日逐点增长率的电力系统短期负荷预测[J]．武汉大学学报，2020, 53(2): 139-144.
CHEN Peiyin，FANG Yanjun Bo．Short-term load forecasting of power system for holiday point-by-point growth rate based on Kalman filtering [J]．Engineering Journal of Wuhan University，2020, 53(2): 139-144.

[8] Rendon-Sanchez J F, De Menezes L M. Structural combination of seasonal exponential smoothing forecasts applied to load forecasting [J]. European Journal of Operational Research, 2019, 275(3): 916-924.

[9] 朱继忠，董瀚江，李盛林，等．数据驱动的综合能源系统负荷预测综述．中国电机工程学报.https://doi.org/10.13334/j.0258-8013.pcsee.202337
ZHU Jizhong, DONG Hanjiang, LI Shenglin, et al. Review of data-driven load forecasting for integrated energy system[J]. Proceedings of the CSEE, https://doi.org/10.13334/j.0258-8013.pcsee.202337.

[10] 牛东晓，谷志红，邢棉，等．基于数据挖掘的 SVM 短期负荷预测方法研究[J]．中国电机工程学报，2006，26(18)：6-12．
NIU Dongxiao，GU Zhihong，XING Mian，et al．Study on forecasting approach to short-term load of SVM based on data mining[J]．Proceedings of the CSEE，2006，26(18)：6-12．

[11] 谷云东,马冬芬,程红超.基于相似数据选取和改进梯度提升决策树的电力负荷预测[J].电力系统及其自动化学报,2019,31(05):64-69.
GU Yundong, MA Dongfen, CHENG Hongchao. Power Load Forecasting Based on Similar-data Selection and Improved Gradient Boosting Decision Tree [J]. Proceedings of the CSU-EPSA, 2019,31(05):64-69.

[12] Dai Y，Zhao P. A hybrid load forecasting model based on support vector machine with intelligent methods for feature selection and parameter optimization[J]. Applied Energy,2020,279:115332.

[13] Massaoudi M ，Refaat S S，Chihi I，et al. A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting [J]. Energy，2021，214：118874.

[14] 张宇帆，艾芊，林琳，等．基于深度长短时记忆网络的区域级超短期负荷预测方法 [J]．电网技术, 2019, 43(6): 1884-1892.
ZHANG Yufan，AI Lin，LIN Lin，et al. A Very Short-term Load Forecasting Method Based on Deep LSTM RNN at Zone Level[J]．Power System Technology，2019, 43(6): 1884-1892．

[15] 张振中,郭傅傲,刘大明,等.基于最大互信息系数和小波分解的多模型集成短期负荷预测[J].计算机应用与软件,2021,38(05):82-87.
ZHANG Zhenzhong, GUO Fuao, LIU Daming, et al. Multi-model integrated short-term load prediction based on maximum mutual information coefficient and wavelet decomposition[J]. Computer Applications and Software, 2021,38(05):82-87(in Chinese).

[16] 刘亚珲,赵倩.基于聚类经验模态分解的 CNN-LSTM 超短期电力负荷预测 [J/OL]. 电网技

术 :1-8[2021-05-21].https://doi.org/10.13335/j.1000-3673.pst.2021.0016.
Liu Yahui , Zhao Qian. Ultra-short-term power load forecasting method based on cluster empirical mode decomposition of CNN-LSTM[J]. Power System Technology,
1-8[2021-05-21].https://doi.org/10.13335/j.1000-3673.pst.2021.0016(in Chinese).
[17] Bento P M R，Pombo J A N，Calado M R A，et al. Optimization of neural network with wavelet transform and improved data selection using bat algorithm for short-term load forecasting[J]. Neurocomputing， 2019，358:53-71.
[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, u. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, p. 6000–6010.
[19] Yi T , Dehghani M , Bahri D , et al. Efficient Transformers: A Survey, arXiv:2009.06732.
[20] Guo S, Lin Y, Wan H, et al. Learning Dynamics and Heterogeneity of Spatial-Temporal Graph Data for Traffic Forecasting[J]. IEEE Transactions on Knowledge and Data Engineering, doi: 10.1109/TKDE.2021.3056502.
[21] Xu M, Dai W, Liu C, et al. Spatial-Temporal Transformer Networks for Traffic Flow Forecasting, arXiv:2001.02908v1.
[22] Sneha Chaudhari, Varun Mithal, Gungor Polatkan, Rohan Ramanath. An Attentive Survey of Attention Models. CoRR abs/1904.02874 (2019)
[23] 朱江行,邹晓松,熊炜,等.基于Prophet与XGBoost混合模型的短期负荷预测[J].现代电力,2021,38(03):325-331.
ZHU Jianghang, ZOU Xiaosong, XIONG Wei, et al. Short-Term Power Load Forecasting Based on Prophet and XGBoost Mixed Model[J]. Modern Electric Power, 2021,38(03):325-331.
[24] Shi X, Yeung D, Machine learning for spatiotemporal sequence forecasting: A survey, CoRR abs/1808.06865.

收稿日期：2021-xx-xx

作者简介
黄飞虎(1990)，男，博士，研究方向为电力市场、电力大数据、时空大数据、深度学习，E-mail：hd808080@126.com；
赵红磊(1981)，男，硕士，研究方向为电网调度、电力市场，E-mail：407936629@qq.com；
弋沛玉(1995)，男，博士研究生，研究方向为时空大数据、电力大数据、深度学习，E-mail：846478748@qq.com；
李沛东(1998)，男，硕士研究生，研究方向为电力市场、电力现货，E-mail：392457608@qq.com；
彭舰(1970)，男，博士，教授，研究方向为大数据挖掘与应用技术、边缘计算、嵌入式系统，E-mail：975905819@qq.com；

（责任编辑：XXX）